

Unsteady pressure patterns discovery from high-frequency sensing in water distribution systems

Lu Xing, Lina Sela*

Department of Civil, Architectural and Environmental Engineering, University of Texas at Austin, Texas, 78712, USA

ARTICLE INFO

Article history:

Received 22 November 2018

Received in revised form

18 March 2019

Accepted 25 March 2019

Available online 28 March 2019

Keywords:

Transients detection

CUSUM

Dynamic time warping

K-means clustering

ABSTRACT

Pressure transients have been identified as one of the major contributing factors in many pipe failures in water distribution systems (WDSs). The behavior of these pressure transients is largely unknown and cannot be fully assessed by numerical simulation or modeling. This study investigates the behavior of pressure transients in WDSs as measured by high-frequency pressure sensors. A Time Series Data Mining (TSDM) approach is proposed to detect and cluster pressure transients to reveal recurrent and consistent patterns. The proposed technique, based on a modified two-sided cumulative sum (CUSUM) algorithm, is used to detect pressure transients. Dynamic Time Warping (DTW) is adopted to measure the similarity between the detected pressure transients, and k-means clustering algorithm is used to discover the characteristic patterns. Several performance scores are suggested to evaluate the quality of the clustering results, including sum of squared error, Silhouette index, and Calinski-Harabaz index. Results demonstrate that the proposed approach is able to reveal consistent and unique patterns across multiple sensing locations. The proposed approach provides a fast and efficient way to discover the hidden information in WDSs by analyzing high-frequency pressure signals from distributed sensors.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Spanning over one million miles across the United States (U.S.), water distribution systems (WDSs) function as complex infrastructure networks to maintain the reliable and safe supply of drinking water (USEPA, 2016). The American Society of Civil Engineers (ASCE) estimates that the aging and deteriorating infrastructures are wasting 14%–18% of the treated water (ASCE, 2017). Due to the relation between pressure, pipe failures, and water loss, pressure management has become one of the most popular management interventions implemented by water utilities in their efforts to reduce pipe failures and water loss (McKenzie and Wegelin, 2009). The primary objective of almost all measures of pressure management constitutes reducing the maximum steady amplitude and the excessive *unsteady variability of the pressure*. While the contributions of the former to pipe failure have been intensively studied (Kabir et al., 2015; Martínez-Codina et al., 2015), there is a lack of quantitative research on the roles that the latter, pressure variability, plays in WDSs (Ghorbanian et al., 2016).

Flow conditions in pipelines can be disrupted by pipe failures (background leakages and bursts), system operations, and demand fluctuations, consequently creating corresponding unsteady changes in water pressure, commonly termed pressure transients or water hammers. These events occur quickly but have the potential to result in the pipe deterioration by introducing extreme pressure variability and intriguing fatigue (Starczewski et al., 2015). Despite the general understanding that pressure variability can contribute to pipe deterioration, limited research has focused on investigating these relations quantitatively (Hoskins and Stoianov, 2014).

Historically, the main reasons for this lack of understanding are the technological constraints, such as the limited availability of high-resolution pressure transmitters to measure the pressure and the computation capacity to process the data. Therefore, most studies about pressure transients were restricted to modeling (Wylie et al., 1993; Ghidaoui et al., 2005) using various of numerical simulation techniques (De Almeida and Koelle, 1992; Karney and McInnis, 1992; Boulous et al., 2005; Wood et al., 2005; Chaudhry, 2014). On the basis of these numerical simulation, various studies have been conducted to investigate drinking water quality due to pressure transients (Ebacher et al., 2012), leak detection (Brunone and Ferrante, 2001; Colombo et al., 2009), pipe deterioration and

* Corresponding author.

E-mail addresses: xinglu@utexas.edu (L. Xing), linasela@utexas.edu (L. Sela).

corrosion assessment (Gong et al., 2012). Unfortunately, without the availability of high-resolution pressure data, the validation of the numerical results has been limited.

Over the past few years, the rapid development of data logging and data mining technologies has relieved some of the aforementioned constraints and made it possible to investigate the pressure transients in WDSs in a more rigorous manner. The emerging applications of low-cost high-frequency transient pressure transmitters (TPTs), introduced to continuously monitor water pressure, directly contribute to the pressure management in WDSs (Allen et al., 2011; Trimble Water Website, 2018; Visenti, 2018). High frequency pressure data have been foremost utilized to validate and calibrate the simulation models by comparing and fitting the numerical results to collected pressure data (Friedman and Friedman, 2004; Ebacher et al., 2010; Meseguer et al., 2014; Rathnayaka et al., 2016). More recently, data-driven approaches have been proposed to detect and localize pipe bursts and leaks (Misiunas et al., 2005; Srirangarajan et al., 2013; Lee et al., 2015b; Wu et al., 2016).

The scope of previous works has been limited due to the typically short duration of sensors deployment and more importantly low sampling frequency. However, pressure transients, reflecting the responses of WDSs to normal and abnormal changes, are manifested as pressure waves traveling in the WDSs with very high velocity in the range of 100–1400 m/s (Ghidaoui et al., 2005). This implies that lower sampling resolutions ($\sim 5 - 15$ min) characteristic of Supervisory Control and Data Acquisition (SCADA) systems prevent from monitoring transient pressures and high resolution pressure data are necessary to capture and investigate the behavior of pressure transients (Srirangarajan et al., 2013). TPTs extend the capabilities of traditional SCADA systems by providing real-time information at a fine spatial-temporal resolution that was previously unavailable. The high-resolution pressure data reveals additional valuable information of WDSs, thus contributing to the improvement of network control, pressure management, water loss control, pipe burst and leakage detection (Puust et al., 2010; Lee et al., 2015a). On the other hand, the enormous volumes of high-resolution pressure data, as utilized in this work, present new challenges in data analysis, information extraction and knowledge discovery. In addition, none of the prior studies considered using the data collected by the TPTs to characterize the typical patterns of pressure transients occurring in WDSs and the corresponding unsteady pressure variability.

Motivated by the availability of high-frequency pressure data collected over an extended period of time, we propose a time series data mining procedure to *detect the pressure changes and discover the patterns of pressure transients from high-resolution pressure data collected by a network of TPTs*. Explicitly, the objectives of this work are to: (a) deploy a network of high-frequency TPTs for an extended period of time at various locations in a WDS; (b) automate the detection of transient pressure changes; (c) automate the classification of transient pressure patterns and pressure intensity; and (d) quantitatively assess and validate the proposed approach.

2. Methods

In this study, time series data mining (TSDM) is identified as a promising approach to explore the behavior of pressure transients in WDSs. TSDM refers to the process of non-trivial extraction of implicit, previously unknown and potentially useful information from time series data (Fayyad et al., 1996). A two-step TSDM approach is proposed for transient detection and pattern discovery based on high-resolution pressure data collected by a network of pressure sensors. In the first step, raw pressure data are pre-processed, from which pressure changes are detected based on a modified cumulative sum control chart (CUSUM) approach. In the

second step, the pressure transients are extracted from the original time series and then clustered based on their similarity; it follows, then, that a prototype is discovered for each cluster to represent the corresponding characteristic pattern. Finally, pressure intensity plots are generated to assess the pressure variability for each characteristic transient. The main steps of the proposed algorithm are described in subsequent sections and depicted in Fig. S1 in the Supporting Information (SI).

2.1. Time series representation

Due to the impulsive noises and large dimensionality presented in the raw pressure time series, the first step in this analysis is to represent the time series in a lower dimensional space. Time series representation is integral in several perspectives (Aghabozorgi et al., 2015): firstly, it reduces the dimension of the original data so that efficiency of the TSDM techniques can be significantly improved (Keogh et al., 2004); secondly, it is capable of eliminating noises, the distracting information in the raw data, thus allowing the processes, e.g., distance measurement and clustering, to focus on the actual functional data while minimizing the bias towards noises (Ratanamahatana and Keogh, 2005). Myriad techniques have been developed for time series representation, including but not limited to Fourier transformations (Agrawal et al., 1993; Keogh et al., 2001), wavelets (Chan and Fu, 1999), symbolic aggregate approximation (SAX) (Lin et al., 2003), and piece-wise aggregate approximation (PAA) (Keogh and Pazzani, 2001). A comprehensive review of time series representation can be found in Aghabozorgi et al. (2015) and Keogh et al. (2004).

In this study, PAA is adopted to reduce the dimensionality of the measured pressure time series due to its simplicity and high computational performance (Keogh et al., 2004). The main idea behind PAA is to reduce the dimensionality of the original data through data integration over time. Given a time series $X = (x_1, x_2, \dots, x_n)$ and the dimension of the transformed space N , we index $(1 \leq N \leq n)$ and assume that N is a factor of n . Then, the time series X of length n can be represented in N space by a vector $\bar{X} = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N)$. The i^{th} element of \bar{X} is calculated by the following equation (Keogh and Pazzani, 2001):

$$\bar{x}_i = \frac{N}{n} \sum_{j=\frac{n}{N}(i-1)+1}^{\frac{n}{N}i} x_j \quad (1)$$

Ultimately, PAA algorithm reduces the time series from n dimensions to N dimensions by dividing the data into N equally sized segments. The resolution of PAA algorithm is then defined by the resolution of the resulting N -dimensional data.

2.2. Change detection

To demonstrate the proposed procedure, it is imperative to define two events used throughout this work:

- **Change:** defined in relation to the preceding data that is occurring faster than the expected rate and greater than the expected amplitude, including positive (raise) and negative change (drop). The analysis in this section pertains to pressure change detection.
- **Transient:** refers to any pressure wave that is short lived. Pressure transients may include several changes. The analysis in the subsequent three sections pertains to pressure transient extraction.

A growing number of techniques have been developed to detect changes in high-frequency pressure signals (Colombo et al., 2009; Xu and Karney, 2017) where discrete wavelet transformation (DWT) (Ferrante et al., 2007; Srirangarajan et al., 2013; Lee et al., 2015b) and CUSUM (Misiunas et al., 2005; Lee et al., 2015b) are the two most widely-used techniques. DWT-based algorithms decompose the one-dimensional pressure signal into different temporal and frequency scales by computing the convolution of the pressure signal with a given wavelet of different period and phase. The decomposition permits the exploration of the temporal variability in the signal at different scales. For example, one can employ lower-scale decompositions for the localization of high-frequency components, or apply higher-scale decompositions for the localization of low-frequency components to retain more global information. As the decomposition level of DWT increases, the noises in the signal are suppressed and the singularities are emphasized (Mallat and Hwang, 1992; Ferrante et al., 2007). This property makes DWT a promising approach for detecting abrupt changes.

However, it does not serve as an appropriate option in our application for several reasons. Firstly, it is possible that the minor pressure changes are emphasized and recognized as significant, thereby leading to false alarms (Lee et al., 2015b). Secondly, the performance of DWT largely depends on the mother wavelet; thus, it is integral to choose a mother wavelet that resembles the natural variability of the data (Khalil and Duchêne, 1999). Consequently, it is difficult to choose an appropriate mother wavelet to resemble all pressure changes, which are essentially different. Thirdly, the parameters in DWT, such as the level and the threshold of the decomposition coefficients, are nonintuitive with vague physical meanings, making it difficult to determine these parameters.

In this work, we rely on the modified CUSUM detection scheme since it exhibits the following characteristics: (a) the scheme has no bias on the change patterns and is able to detect changes of any shape as long as it occurs faster than expected rate and greater than expected amplitude; and (b) the scheme is highly interpretable, i.e., involving as few parameters as possible that have clear physical meanings and can be determined using design principles.

The traditional two-sided CUSUM was proposed by Page (1954) as two repeated uses of sequential probability tests for detecting an increase, c^+ , and the decrease, c^- , in the mean of the signal (Gustafsson, 2000). The CUSUM calculation is initialized by setting $c^+(0) = c^-(0) = 0$. At each time step, t , CUSUM tracks the characteristics of the changes in the pressure signal, i.e., rate and magnitude, and compares these characteristics with control limits, i.e., *drift* and *threshold*, as follows:

$$\begin{cases} dp(t) = p(t) - p(t-1) \\ c^+(t) = c^+(t-1) + dp(t) - \text{drift} \\ c^-(t) = c^-(t-1) - dp(t) - \text{drift} \end{cases} \quad (2)$$

where $p(t)$ is the pressure data collected at time t , $dp(t)$ is the pressure variation between time t and $t-1$, and *drift* represents the minimum countable change rate. If either $c^+(t)$ and $c^-(t)$ are greater than zero, indicating that the pressure change within one time step is larger than *drift*, then either $c^+(t)$ or $c^-(t)$ will increase, respectively. Otherwise, when the pressure change per time step is smaller than the minimum countable change rate, *drift*, $c^+(t)$ or $c^-(t)$ will be set to zero, and the corresponding time will be labeled as a candidate start time ($t'_s = t$) of the prospective change. Then, as the changes accumulate, when either $c^+(t)$ or $c^-(t)$ exceed the specific control limit *threshold*, an alarm is raised, marking the detection of an abrupt change; it follows, then, that the alarm ($t_a = t$) and the start ($t_s = t'_s$) times are recorded.

The physical meaning of the two parameters is clear: *drift*, relates to the control limit of the pressure changes within one time

step, and *threshold*, represents the control limit of the accumulative amplitude of change. However, quantitatively, since c^+ or c^- are subtracted by *drift* every time step, they are significantly smaller than the actual pressure change between the start point and the alarm point. As a result, *threshold* does not control the actual limit of the pressure change amplitude, but a considerably smaller value instead, especially when *drift* is a large number. To resolve this problem, c_{real}^+ and c_{real}^- are introduced to record the actual amplitude of the pressure change, as defined below:

$$\begin{cases} c_{\text{real}}^+(t) = c_{\text{real}}^+(t-1) + dp(t) \\ c_{\text{real}}^-(t) = c_{\text{real}}^-(t-1) - dp(t) \end{cases} \quad (3)$$

As before, per-time-step pressure changes that are smaller than *drift* are eliminated by setting c^+ , c^- , c_{real}^+ and c_{real}^- to zeros and the candidate start time (t'_s) is labeled based on the positiveness of c^+ and c^- . The major modification is that c_{real}^+ and c_{real}^- , instead of c^+ and c^- , are adopted as the test statistics to test whether the amplitude of pressure changes exceed the *threshold*. In other words, the alarm timestamps (t_a) are recorded based on the relative relation between c_{real}^+ or c_{real}^- and *threshold*. Thereby, the *threshold* can represent the exact amplitude limit for the pressure changes. Algorithm 1 summarizes the modified CUSUM algorithm to detect abrupt changes in the pressure signal.

Algorithm 1 (Detecting pressure changes)

- 1: **Input:** pressure time-series $p(t)$, *drift*, *threshold*
 - 2: **Output:** vector of start (t_s) and alarm (t_a) times of pressure changes
 - 3: **Initialize:** $c^+(0) = c^-(0) = c_{\text{real}}^+(0) = c_{\text{real}}^-(0) = 0$
 - 4: **While** $t \leq T$ **do:**
 - (i) Compute $c^+(t)$, $c^-(t)$ using Equation 2
 - (ii) Compute $c_{\text{real}}^+(t)$, $c_{\text{real}}^-(t)$ using Equation 3
 - (iii) **If** $c^+(t) \leq 0$ or $c^-(t) \leq 0$ **do:**

$$c^+(t) = c^-(t) = c_{\text{real}}^+(t) = c_{\text{real}}^-(t) = 0$$
Record candidate change start time $t'_s = t$
 - (iv) **If** $c_{\text{real}}^+(t) \geq \text{threshold}$ or $c_{\text{real}}^-(t) \geq \text{threshold}$ **do:**
Record change alarm time $t_a = t$ and the corresponding start time $t_s = t'_s$
 - (v) **Advance** to next time step: $t = t + 1$
-

To detect the end time of the change (t_e), the same procedure is executed on the reverse time series. Then, (t_s) and (t_e) respectively denote the starting and ending time of the change detected at alarm time (t_a). In the results section, we will show how c^+ , c^- , c_{real}^+ , c_{real}^- are updated as well as how *drift* and *threshold* are selected based on the events of interest. At the end of the change detection process, we are able to automatically detect rapid changes in the pressure signal.

2.3. Transient extraction

A collection of subsequent pressure changes forms a pressure transient event (as defined in the previous section). After individual changes are detected, pressure data of a certain duration, i.e., window size, are collected for further analysis. The analysis window is required to contain sufficient amount of information to represent a complete transient event, including several sequential pressure changes. Moreover, as a pressure transient dissipates, by the end of an analysis window, we expect the unsteady pressure to stabilize without significant oscillations. To identify the characteristic window size, we define the absolute pressure difference between the last two points of the analysis window as:

$$\delta p = |p_{-1} - p_{-2}| \quad (4)$$

where the pressure difference at the end of the transient should approach zero, $\delta p \rightarrow 0$. In the results section, we provide details on

selecting the characteristic window size for pressure transients based on the mean and variance of δp .

2.4. Distance measure

Distance measure, quantitatively illustrating the level of similarity between sequences, is of fundamental importance to a variety of time series mining tasks, including clustering. Time series distance measures could be classified into three categories: model based metric, non-elastic metric, and elastic metric (Aghabozorgi et al., 2015). In model based metrics, such as ARMA process (Kalpakis et al., 2001; Xiong and Yeung, 2002), the original data are first fitted into a certain statistical model, and then distances are measured based on the parameters of the fitted model. Non-elastic metrics, such as Euclidean distance, operate in the time domain and measure the distance by comparing the values at each time step. In contrast, elastic metrics compensate for potential temporal misalignment through some elastic adjustment (Lines and Bagnall, 2015). Thereby, if the occurrence of the patterns in time is of concern, elastic metrics, such as dynamic time warping (DTW) (Berndt and Clifford, 1994), allow robust time distance calculation, while non-elastic metric can be extremely brittle because of their sensitive nature to distortions in time axis (Chu et al., 2002). Therefore, the DTW elastic distance metric is adopted in this study.

The DTW algorithm, as the most widely-used elastic metric, can be briefly introduced as follows (Keogh and Pazzani, 2001): to align two given time series $X_1 = (x_{11}, x_{12}, \dots, x_{1n})$ and $X_2 = (x_{21}, x_{22}, \dots, x_{2m})$ using DTW, an $n \times m$ Euclidean distance matrix (d) is first constructed, with $d(i, j)$ being the Euclidean distance between points x_{1i} and x_{2j} . Predicated on the Euclidean distance matrix, the cumulative distance matrix (γ) and the corresponding wrapping path, i.e. the path minimizing the warping cost between the two signals, are calculated using dynamic programming as shown in Equation (5). The control variable w restricts the bandwidth of the search path and thus limits the allowed temporal misalignment between the two signals (Fu et al., 2008).

$$\gamma(i, j) = d(i, j) + \min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\}, \quad (5)$$

$$j - w \leq i < j + w$$

Intuitively, Equation (5) recursively computes the least-cost distance between each pair of points (i, j) while considering the pairwise Euclidean distance $d(i, j)$ plus the cumulative distances it takes to get to point (i, j) from $(0, 0)$ and allowing for i and j to take any values, i.e. not restricting the path to be $i = j$ as in the Euclidean distance. Then, $\gamma(n, m)$ represents the DTW distance. Fig. 1 schematically illustrates the DTW distance measure with a limited bandwidth w , where each cell in the shaded matrix represents the cumulative distance $\gamma(i, j)$, and the solid white line represents the warping path with minimum cost between the two signal X_1 and X_2 . In our application, different pressure signals are expected to exhibit temporal misalignment (e.g. starting time of the signal) as well as different shapes and magnitudes. The DTW distance measure calculates the best match that accounts for the temporal shift, elongation, and compression of the signals in time domain. Hence, the DTW distance allows capturing the general shape of the signal, rather than the perfect matching. Once defined, the DTW distance measure can be used to evaluate the similarity between different pressure transient signals extracted in previous step.

2.5. Pattern discovery

The objective of pattern discovery is to identify an unknown subset of signals that occur frequently in a dataset and then make high level summaries of the massive data set accordingly. In this

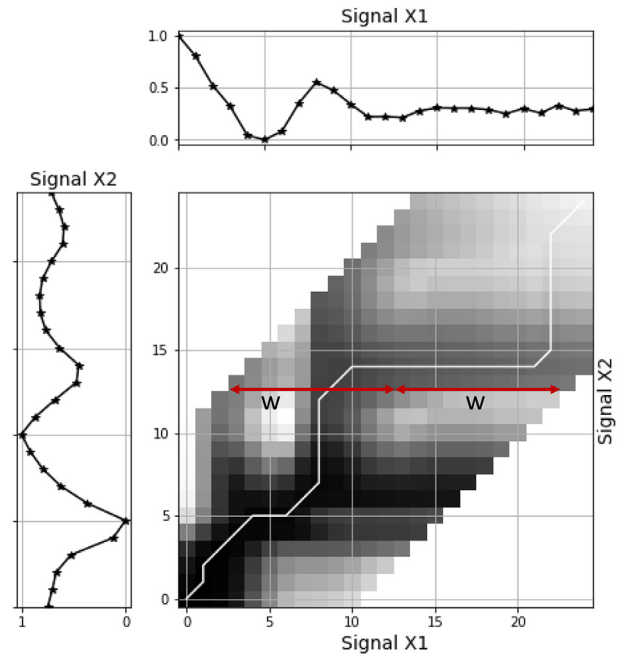


Fig. 1. DTW distance between signal X_1 and signal X_2 with limited bandwidth w , the shaded matrix represents the cumulative distance γ ; the solid line represents the warping path with minimum cost. The darker elements represent lower cumulative distances.

study, pressure transient signals have already been identified from the complete time series based on the detected changes and extracted transients in previous sections. The next step is to classify the characteristic patterns from the set of extracted pressure transients. For this objective, clustering is the most commonly used technique (Fu, 2011) as a solution for classifying enormous data without early knowledge about the classes (Aghabozorgi et al., 2015). The problem of time-series clustering is defined as follows (Aghabozorgi et al., 2015). Given a dataset (D) composed of n time-series signals $D = \{X_1, X_2, \dots, X_n\}$, the process of unsupervised partitioning of D into $C = \{C_1, C_2, \dots, C_k\}$, in such a way that homogeneous time-series are grouped together based on a certain similarity measure, is called time-series clustering. Then, C_i is called a cluster, where $D = \bigcup_{i=1}^k C_i$ and $C_i \cap C_j = \emptyset$ for $i \neq j$.

Various algorithms have been developed to cluster time series data, including k-means, k-medoids, agglomerative hierarchical, self-organizing maps (Liao, 2005) in myriad domains (Tran and Wagner, 2002; Lee et al., 2006; Fu, 2011). Among all clustering algorithms, k-means clustering is the most popular one because of its computational efficiency and performance (Han et al., 2001; Wu et al., 2008). The procedure of k-means clustering can be briefly explained as follows. Initially, k objects are randomly selected, each of which represents an initial center or mean of a cluster. Secondly, each object in the data set is assigned to the closest cluster based on the distance measured between an object in the dataset and the center of the cluster mean. Thirdly, the new mean for each cluster is recalculated based on newly classified objects. The process iterates until some criteria function converges or no improvement/change is achieved. The DTW distance measure is used to cluster pressure transients using the k-means algorithm.

The objective of this study is to reveal recurrent and consistent patterns of the pressure transients; therefore, a prototype is required to represent each cluster. Herein, the prototype of the i^{th} cluster is defined as its medoid (M_i), whose DTW distance to all other transients in that cluster is minimal.

3. Results and discussion

3.1. Data preparation

The methodology proposed here is tested on real data, collected by high-frequency transient pressure transmitters distributed in a large water utility. Each TPT unit includes a pressure sensor taking 64 samples per second and a remote telemetry unit, which transmits the data to a server. The TPT unit is mounted on a fire hydrant as shown in Fig. 2(a). An example of a pressure signal recorded by a TPT during one day is shown in Fig. 2(b). The dataset available in this study is high-resolution pressure data collected by TPTs from October 2017 to August 2018 and is provided in the SI.

As the preprocessing step, PAA segmentation is applied to the raw pressure data to reduce dimension and eliminate noises. To determine the dimensionality reduction, several resolutions are tested. The effect of dimensionality reduction on the number of detected pressure changes per day are compared in Fig. S2 of the SI. As the PAA resolution gets coarser, the number of detected changes decreases because of the averaging nature of PAA algorithm. Additionally, the finer the PAA resolution, the more computation effort is required for the proposed algorithms. Therefore, there is a trade-off between efficiency and accuracy in the choice of PAA resolutions. In this study, PAA resolution is chosen to be 10s, as this provides the greatest gain in computational efficiency and only small compromise in accuracy.

3.2. Change detection

The pressure changes detected by CUSUM algorithm are characterized by the start time (t_s) and the pressure at start time (p_{t_s}) as well as the end time (t_e) and the pressure at end time (p_{t_e}). The duration (T) of the pressure change is then defined by $T = t_e - t_s$. This work focuses on significant and abrupt pressure changes satisfying the following requirements, based on which the parameters in CUSUM algorithm are chosen:

1. The pressure change is significant enough that the amplitude (Δp) is greater than 10 psi:

$$\Delta p = p_{t_e} - p_{t_s} > 10 \text{ psi} \quad (6)$$

Accordingly, the *threshold* parameter of CUSUM algorithm is set to 10 psi.

2. The pressure change is abrupt enough that the rate of change is greater than 0.1 psi/s:

$$\frac{dp}{dt} = \frac{p_{n+1} - p_n}{t_{n+1} - t_n} > 0.1 \text{ psi/s} \quad (7)$$

where dt is the PAA resolution. In this study, dt equals 10 s because the PAA resolution is set to be 10 s. Therefore, the *drift* parameter is set to $0.1 \text{ psi/s} \times 10 \text{ s} = 1 \text{ psi}$.

With the parameters discussed above, Fig. 2(b) shows CUSUM results of the pressure data collected from TPT#1 in the network within a typical day, when 11 pressure changes are detected. The right arrows represent the starting point (t_s, p_{t_s}) and the left arrows represent the ending points (t_e, p_{t_e}) of the detected changes. Fig. 2(c) illustrates the first two detected pressure change: a pressure raise followed by a pressure drop, the combination of which constitutes a pressure transient. The corresponding CUSUM results for the first two detected changes are then presented in Fig. 2(d). When the CUSUM value (c_{real}^+ or c_{real}^-) exceeds the *threshold*, represented by the red dashed line, an alarm will be raised, illustrated by the red dots; consequently, the start and end point will be recorded.

The CUSUM algorithm is then applied to pressure data collected by TPT #1 during the entire time period, where 1314 pressure changes were detected. The historical analysis of the pressure changes is represented in Fig. S3 showing the distribution of the pressure change amplitude (Δp) and pressure change duration (T). Notably, the amplitude distribution is not symmetric about 0, instead it is skewed towards the left, indicating that more negative

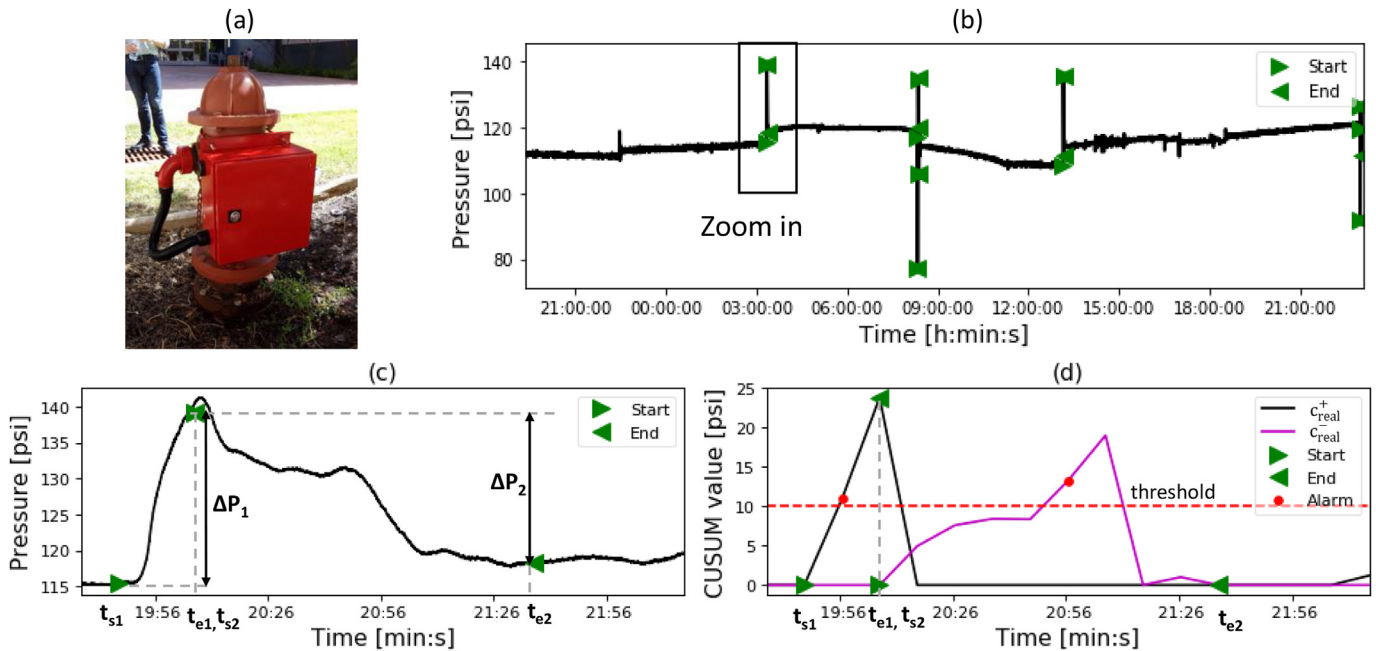


Fig. 2. Change detection by applying CUSUM algorithm, where right arrows, left arrows, and dots denote start, end, and alarm time respectively: (a) TPT#1 installed on a fire hydrant; (b) daily pressure data and the detected changes; (c) the zoom-in view of one pressure signal with start (right arrow) and end (left arrow) points; (d) the zoom-in view of the CUSUM results.

changes (drop) occurred than their positive counterparts (raise) during this time period. The pressure change duration varies from 10 s to 150 s with an average duration of 60 s.

3.3. Transient extraction

The analysis window represents the duration of the pressure transients and should encompass all necessary information to represent a complete pressure transient event, typically comprised of several pressure changes. For example, the transient event shown in Fig. 2(c) constitutes two sequential pressure changes: one positive change and one negative change. Subsequently, on the basis of detected changes in the previous section, several window sizes for transient extraction are tested and the absolute pressure difference between last two points of the analysis window (δp) is calculated for the entire data. Fig. 3 shows the mean and variance of the δp distribution. It can be noticed that the mean and variance decrease as the window size increases from 1 min to 5 min and stabilize after that. Therefore, 5 min, as the knee point, is chosen to be the window size to extract transient events and perform the following analysis. Based on the 1314 detected pressure changes, 586 5-min transient events are extracted from the original data.

3.4. Similarity search

In terms of similarity search, our interests lie in the similarity in patterns instead of the absolute amplitude; therefore, the extracted pressure transients are normalized. DTW distances between every pair of the 586 normalized transient events are calculated with bandwidth $w = 5$, allowing a maximum temporal misalignment of 50 s, which is shorter than the mean duration of a typical pressure change (60 s) (as shown in Fig. S3), thereby avoiding possible excessive misalignment. Consequently, the pairwise DTW distances comprise a 586×586 distance matrix, of which the $(i, j)^{th}$ element represents the DTW distance between i^{th} and j^{th} transient event. The distance matrix between each pair of detected pressure transients based on their chronological appearance is shown in Fig. 4(a). The darker the $(i, j)^{th}$ pixel, the lower the value of $(i, j)^{th}$ element is, and the more similar the i^{th} and j^{th} transient events are.

On the basis of DTW distance matrix, the normalized pressure transients can then be mined by k-means clustering. Since in k-means algorithm the number of clusters has to be predefined, we experiment with different number of clusters. The results show that three clusters are most informative in terms of identifying and distinguishing the patterns for the given dataset. The evaluation of the number of clusters is demonstrated in the next section. After

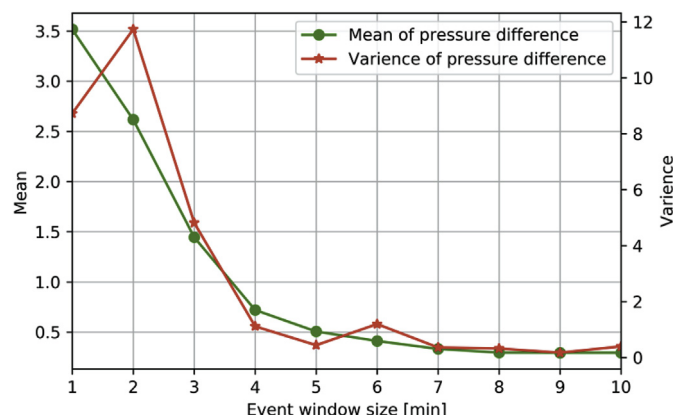


Fig. 3. Statistical description of pressure difference as a function of window size.

clustering, the transient events are sorted based on their cluster labels and the corresponding distance matrix is shown in Fig. 4(b). Compared with the distance matrix before clustering as in Fig. 4(a), the matrix after clustering is generally better organized. Specifically, three distinguished partitions can be identified, corresponding to the three identified clusters. In addition, the diagonal block matrices take lower values than the off-diagonal ones, indicating that the within-cluster distances are smaller than the between-cluster distances.

Fig. 5 shows the clustering results and the identified prototype patterns for each cluster. Explicitly, Cluster 0 includes 212 down-surge pressure transients, where pressure drops at the beginning of the transient event, fluctuates and finally stabilizes. If we define a pair of pressure increase and decrease as a cycle, Pattern 0 constitutes a major cycle and a minor cycle. However, Cluster 1 comprises 251 up-surge transient events, characterized by starting with pressure raise and following by a pressure drop. Only one major cycle exists in this pattern and the pressure stabilizes at a higher pressure level than the original. In Cluster 2, which is made up of 123 transient events, the representative pattern shares some similarity with Pattern 1, but differs in that the pressure of Pattern 2 returns to original level when it stabilizes.

Results for data collected at a different location, TPT#2, are shown in Fig. S4, where four distinguishing patterns are identified. Significant differences can be noticed between the pressure patterns discovered in different stations. Several reasons can contribute to these differences: (1) the relative location of the TPT to the pumping station in the pressure zone; (2) the pumping operation routine in the corresponding pumping station; and (3) the characteristics of the piped network. These differences are instructive in understanding the relation between the transient patterns and pump operations or piped network characteristics, and, in turn, their effects on pipe condition. Rapid changes in flow generate pressure waves that travel through the pipelines. The pressure signals undergo reflection and attenuation, but largely maintain their shape throughout the system (Lee et al., 2015a). The internal characteristics of the pipelines, such as material, diameter, and friction, as well as external characteristics such as leaks and blockages affect transient response and create changes to the shape of the pressure wave (Xu and Karney, 2017). For example, the occurrence of a leak releases some amount of flow through the orifice, thus providing partial transient protection for the system and modifying the magnitude of the transient pressure wave (Xu and Karney, 2017). By analyzing the pressure signals in the time and/or frequency domain system uncertain characteristics and faults can be detected. A variety of transient reflection, damping, and response technique as well as inverse optimization techniques exist for identifying pipeline characteristics and/or faults by tracing the changes in the time of arrival, magnitude, and shape of the pressure transients (Colombo et al., 2009; Chorbani et al., 2016).

3.5. Clustering results analysis

Numerical criteria to evaluate the performance of the clustering algorithm can be classified into two categories: external index, which is appropriate when the labels of ground truth are given, and internal index, which is appropriate when ground truth is unknown (Aghabozorgi et al., 2015). In this study, we do not know in advance the prototype of each cluster and which cluster each signal should belong to. Therefore, only internal index is applicable. We apply three internal indexes to evaluate the performance of the cluster results and determine the optimal number of clusters (k): (1) Silhouette Coefficient – a measure of how well a signal X_i is assigned to its cluster (Rousseeuw, 1987), (2) Sum of Squared Error (SSE) – describes the coherence of the given clusters (Han et al.,

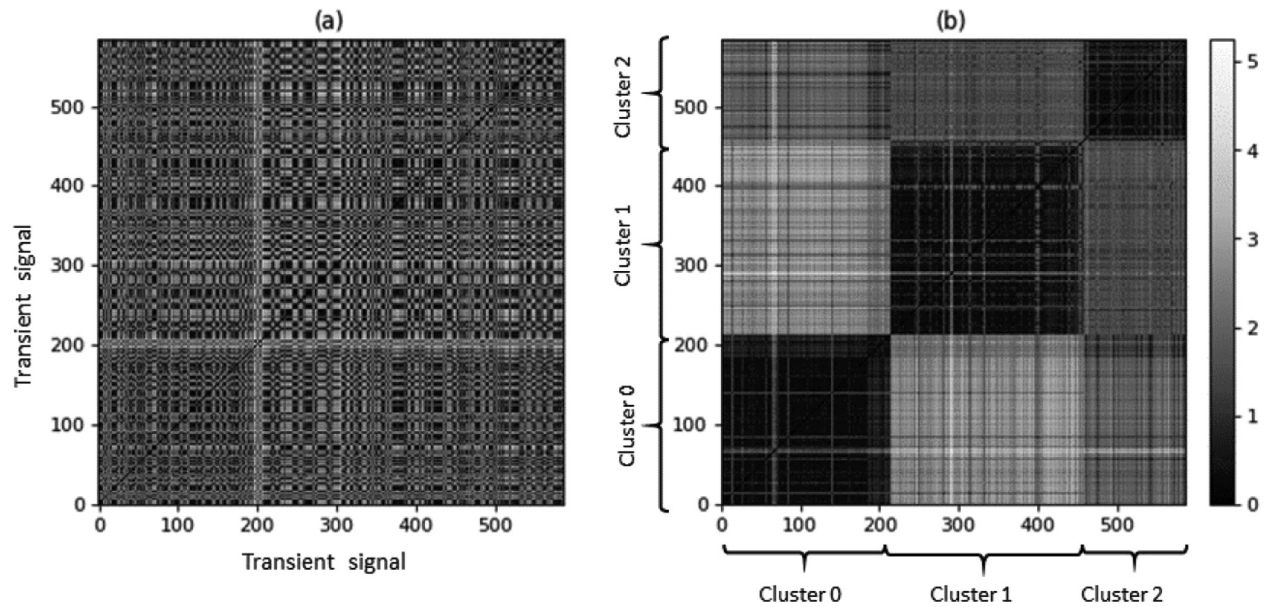


Fig. 4. DTW distance between the pressure transients before (a) and after clustering (b).

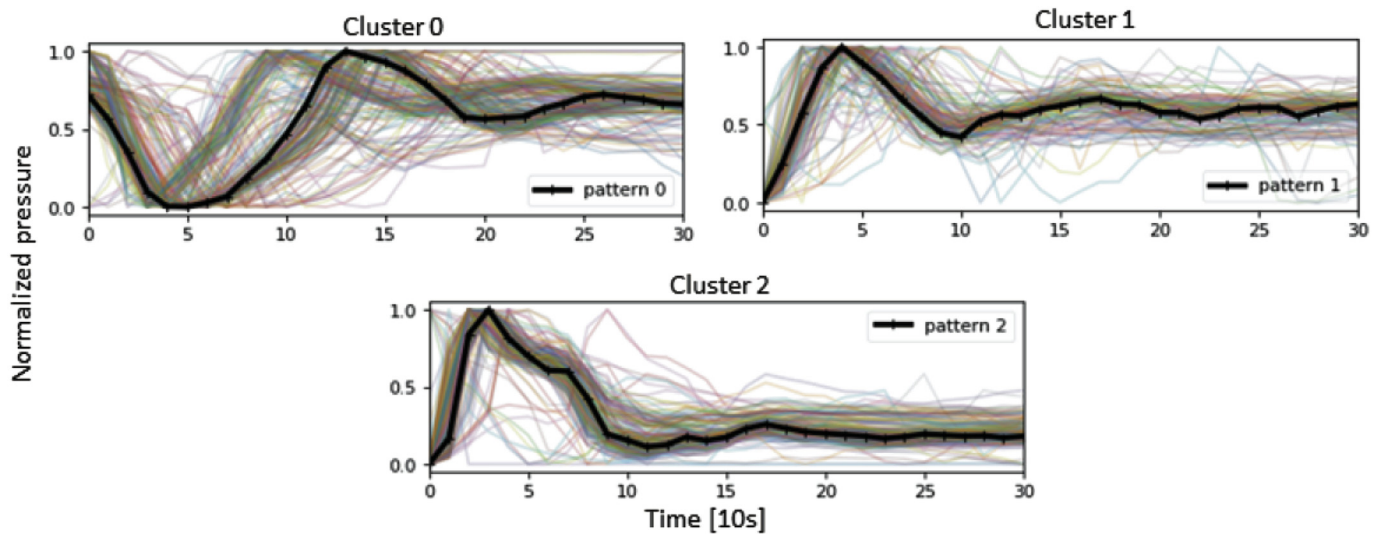


Fig. 5. Clusters and their prototypes at TPT #1.

2011), and (3) Calinski-Harabaz (CH) Index – defined as the ratio between the within-cluster dispersion and the between-cluster dispersion (Caliński and Harabasz, 1974). Detailed explanation of the three indexes can be found in the SI.

Fig. 6 shows Silhouette coefficient, SSE, and Calinski-Harabaz index as a function of the number of clusters for pressure transients recorded at TPT #1. Both averaged Silhouette coefficient and Calinski-Harabaz index peak at three clusters, which is also the knee point for SSE curve. Therefore, cluster number of three is identified as the most informative setting for the given pressure data, as shown in Fig. 5. Similar analysis was conducted for TPT#2, which resulted in four significant transient clusters (see Figs. S6 and S7).

3.6. Pressure intensity

Pressure transients, as shown in this work, are intrinsically

cyclical due to the inherent diurnal pumping operations and demand patterns, and may impose excessive internal stress on the pipelines. Cyclic load, or fatigue, is the most predominant mechanical force that can substantially contribute to pipe failures (Schijve, 2009; Yu et al., 2016). Fatigue failure is caused by repeated alternating stresses of an intensity considerably below the normal strength and it strongly depends on the frequency, intensity, and shape of the stress cycles (Cui, 2002; Schijve, 2009; Rajani and Kleiner, 2010). Having identified the shapes of the characteristic transients, we can evaluate the frequency and intensity of each type of transient event. Pressure intensity is calculated as the absolute difference between the minimum and maximum pressures in a given transient event, and frequency is the number of times a given range of pressure intensity was observed. Fig. 7 shows the cumulative probability distributions of pressure intensity for the three characteristic patterns identified in TPT #1. Upsurge patterns (solid gray and black lines) show more frequent low-intensity transient

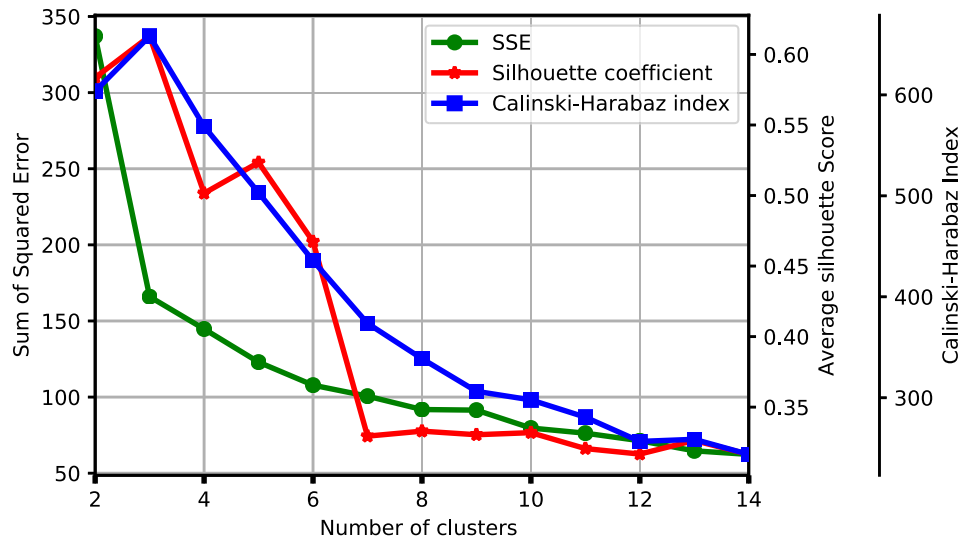


Fig. 6. The Silhouette coefficient, SSE, and Calinski-Harabaz index measured with different cluster number at TPT#1.

and less frequent high-intensity transients. Contrary, the down-surge patterns (dashed black line) show more frequent high-intensity transients. Similar results for TPT #2 are shown in Fig. S8 in the SI.

The frequency, intensity, and shape of the transient pressure patterns discovered and extracted by the proposed method can be applied to improve the work-flow of the calibration of numerical transient models, which cannot be calibrated nor validated without comparison with data. For example, in transient hydraulic analysis, several pipeline parameters are required, including wave speed and friction factor, the evaluation of which is not a trivial task. With the availability of pressure data collected by pressure sensors, it is possible that these parameters can be calibrated more rigorously using inverse optimization techniques. However, if the calibration is performed by minimizing the difference of observed raw pressure and simulated pressure at each time step, it is highly possible that the noises in the observed data would be exaggerated. To resolve this problem, the shape, intensity, and frequency of the patterns can be used as dynamic characteristics of transient pressure to calibrate the numerical models in a more effective and efficient

manner. In the context of condition assessment, previous studies have shown that pipe failures are not only related by the minimum and maximum pressure as traditionally identified from transient hydraulic analysis, but also highly correlated with the shape of pressure transient patterns including duration, frequency, amplitude, as well as sequence of cycles (Yu et al., 2016; Zhao et al., 2016, 2017). However, in water distribution systems the impact and specific contributing features of pressure transients on pipe failures are not yet clear. Therefore, the proposed methodology provides a solid foundation for further studying the cause-effect relationship between pressure transient patterns and pipe failures, thereby contributing to a more comprehensive risk assessment.

4. Conclusion

In this paper, we propose a TSDM approach to investigate pressure variability in WDS by extracting the shape, intensity, and frequency from high-frequency pressure sensors. The CUSUM algorithm was modified to directly relate parameters (i.e., *threshold* and *drift*) to the physical characteristics of pressure changes (i.e., amplitude and gradient). The modified CUSUM change detection algorithm, together with DTW distance measure and k-means clustering technique, was applied to detect and classify the pressure transient events into clusters, for which the representative patterns were then discovered. Several performance scores were suggested to investigate the optimal number of clusters and evaluate the quality of the clustering results. The proposed procedure was demonstrated using pressure data collected by two TPTs located at different pressure zones, where different transient patterns were discovered. The example applications have shown TSDM to be a powerful tool for transient pressure analysis, so as to reveal consistent and unique patterns across multiple sensing locations in WDSs. To maximize the full benefits of transient pressure analysis, future research should be devoted to identifying the source of the transients and its effect on pipe condition and failures.

Acknowledgement

The authors would like to thank Austin Water for providing guidance and support throughout this study. This work was supported by the University of Texas at Austin Startup Grant and the Cooperative Agreement No. 83595001 awarded by the U.S.

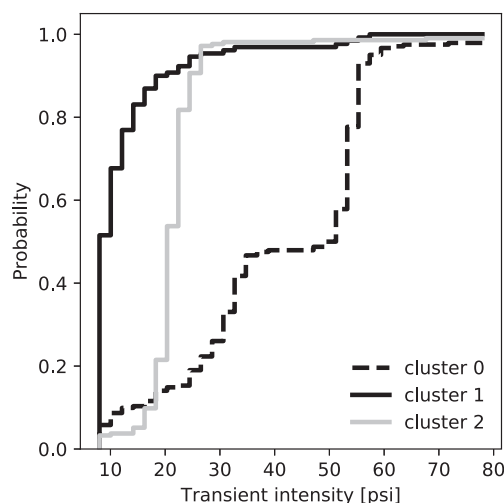


Fig. 7. Cumulative probability distribution of pressure intensity for different clusters at TPT #1: downsurges (dashed) and upsurges (solid).

Environmental Protection Agency to The University of Texas at Austin. It has not been formally reviewed by EPA. The views expressed in this document are solely those of the authors and do not necessarily reflect those of the Agency. EPA does not endorse any products or commercial services mentioned in this publication.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.watres.2019.03.051>.

References

- Aghabozorgi, S., Shirkhorshidi, A.S., Wah, T.Y., 2015. Time-series clustering—a decade review. *Inf. Syst.* 53, 16–38.
- Agrawal, R., Faloutsos, C., Swami, A., 1993. Efficient similarity search in sequence databases. In: *International Conference on Foundations of Data Organization and Algorithms*. Springer, pp. 69–84.
- Allen, M., Prels, A., Iqbal, M., Srirangarajan, S., Lim, H.B., Glrod, L., Whittle, A.J., 2011. Real-time in-network distribution system monitoring to improve operational efficiency. *J. Am. Water Work. Assoc.* 103 (7), 63–75.
- ASCE, 2017. Infrastructure Report Card. Technical Report. ASCE, Reston, VA, 2017.
- Berndt, D.J., Clifford, J., 1994. Using dynamic time warping to find patterns in time series. In: *KDD Workshop*. Seattle, WA, 10, pp. 359–370.
- Boulos, P.F., Karney, B.W., Wood, D.J., Lingireddy, S., 2005. Hydraulic transient guidelines for protecting water distribution systems. *J. Am. Water Work. Assoc.* 97 (5), 111–124.
- Brunone, B., Ferrante, M., 2001. Detecting leaks in pressurised pipes by means of transients. *J. Hydraul. Res.* 39 (5), 539–547.
- Caliński, T., Harabasz, J., 1974. A dendrite method for cluster analysis. *Commun. Stat. Theor. Methods* 3 (1), 1–27.
- Chan, K.P., Fu, W.C., 1999. Efficient time series matching by wavelets. In: *ICDE*. IEEE, p. 126.
- Chaudhry, M.H., 2014. Transient-flow equations. In: *Applied Hydraulic Transients*. Springer New York, New York, NY, pp. 35–64.
- Chu, S., Keogh, E., Hart, D., Pazzani, M., 2002. Iterative deepening dynamic time warping for time series. In: *Proceedings of the 2002 SIAM International Conference on Data Mining*. SIAM, pp. 195–212.
- Colombo, A.F., Lee, P., Karney, B.W., 2009. A selective literature review of transient-based leak detection methods. *Journal of hydro-environment research* 2 (4), 212–227.
- Cui, W., 2002. A state-of-the-art review on fatigue life prediction methods for metal structures. *J. Mar. Sci. Technol.* 7 (1), 43–56.
- De Almeida, A.B., Koelle, E., 1992. *Fluid Transients in Pipe Networks*. Elsevier Applied Science, Windsor, United Kingdom.
- Ebacher, G., Besner, M., Clément, B., Prévost, M., 2012. Sensitivity analysis of some critical factors affecting simulated intrusion volumes during a low pressure transient event in a full-scale water distribution system. *Water Res.* 46 (13), 4017–4030.
- Ebacher, G., Besner, M.C., Lavoie, J., Jung, B., Karney, B., Prévost, M., 2010. Transient modeling of a full-scale distribution system: comparison with field data. *J. Water Resour. Plan. Manag.* 137 (2), 173–182.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., 1996. From data mining to knowledge discovery in databases. *AI Mag.* 17 (3), 37.
- Ferrante, M., Brunone, B., Meniconi, S., 2007. Wavelets for the analysis of transient pressure signals for leak detection. *J. Hydraul. Eng.* 133 (11), 1274–1282.
- Friedman, M., Friedman, M.J., 2004. *Verification and Control of Pressure Transients and Intrusion in Distribution Systems*. AWWA Research Foundation and US Environmental Protection Agency.
- Fu, A.W.C., Keogh, E., Lau, L.Y., Ratanamahatana, C.A., Wong, R.C.W., 2008. Scaling and time warping in time series querying. *The VLDB Journal/The International Journal on Very Large Data Bases* 17 (4), 899–921.
- Fu, T.C., 2011. A review on time series data mining. *Eng. Appl. Artif. Intell.* 24 (1), 164–181.
- Ghidaoui, M.S., Zhao, M., McInnis, D.A., Axworthy, D.H., 2005. A review of water hammer theory and practice. *Appl. Mech. Rev.* 58 (1), 49–76.
- Ghorbanian, V., Karney, B., Guo, Y., 2016. Pressure standards in water distribution systems: reflection on current practice with consideration of some unresolved issues. *J. Water Resour. Plan. Manag.* 142 (8), 04016023.
- Gong, J., Simpson, A.R., Lambert, M.F., Zecchin, A.C., Kim, Y.I., Tijsseling, A.S., 2012. Detection of distributed deterioration in single pipes using transient reflections. *J. Pipeline Syst. Eng. Pract.* 4 (1), 32–40.
- Gustafsson, F., 2000. *Adaptive Filtering and Change Detection*, 1. Citeseer.
- Han, J., Kamber, M., Tung, A.K., 2001. Spatial Clustering Methods in Data Mining. *Geographic data mining and knowledge discovery* 188–217.
- Han, J., Pei, J., Kamber, M., 2011. *Data Mining: Concepts and Techniques*. Elsevier.
- Hoskins, A., Stoianov, I., Infransense, 2014. A distributed system for the continuous analysis of hydraulic transients. *Procedia Engineering* 70, 823–832.
- Kabir, G., Tesfamariam, S., Francisque, A., Sadiq, R., 2015. Evaluating risk of water mains failure using a bayesian belief network model. *Eur. J. Oper. Res.* 240 (1), 220–234.
- Kalpakakis, K., Gada, D., Puttagunta, V., 2001. Distance measures for effective clustering of arima time-series. In: *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*. IEEE, pp. 273–280.
- Karney, B.W., McInnis, D., 1992. Efficient calculation of transient flow in simple pipe networks. *J. Hydraul. Eng.* 118 (7), 1014–1030.
- Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S., 2001. Dimensionality reduction for fast similarity search in large time series databases. *Knowl. Inf. Syst.* 3 (3), 263–286.
- Keogh, E., Chu, S., Hart, D., Pazzani, M., 2004. Segmenting time series: a survey and novel approach. In: *Data Mining in Time Series Databases*. World Scientific, pp. 1–21.
- Keogh, E.J., Pazzani, M.J., 2001. Derivative dynamic time warping. In: *Proceedings of the 2001 SIAM International Conference on Data Mining*. SIAM, pp. 1–11.
- Khalil, M., Duchène, J., 1999. Detection and classification of multiple events in piecewise stationary signals: comparison between autoregressive and multi-scale approaches. *Signal Process.* 75 (3), 239–251.
- Lee, C.H.L., Liu, A., Chen, W.S., 2006. Pattern discovery of fuzzy time series for financial prediction. *IEEE Trans. Knowl. Data Eng.* 18 (5), 613–625.
- Lee, P.J., Duan, H.F., Tuck, J., Ghidaoui, M., 2015a. Numerical and experimental study on the effect of signal bandwidth on pipe assessment using fluid transients. *J. Hydraul. Eng.* 141 (2), 04014074.
- Lee, S.J., Lee, G., Suh, J.C., Lee, J.M., 2015b. Online burst detection and location of water distribution systems and its practical applications. *J. Water Resour. Plan. Manag.* 142 (1), 04015033.
- Liao, T.W., 2005. Clustering of time series data survey. *Pattern Recogn.* 38 (11), 1857–1874.
- Lin, J., Keogh, E., Lonardi, S., Chiu, B., 2003. A symbolic representation of time series, with implications for streaming algorithms. In: *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*. ACM, pp. 2–11.
- Lines, J., Bagnall, A., 2015. Time series classification with ensembles of elastic distance measures. *Data Min. Knowl. Discov.* 29 (3), 565–592.
- Mallat, S., Hwang, W.L., 1992. Singularity detection and processing with wavelets. *IEEE Trans. Inf. Theory* 38 (2), 617–643.
- Martínez-Codina, A., Cueto-Felgueroso, L., Castillo, M., Garrote, L., 2015. Use of pressure management to reduce the probability of pipe breaks: a bayesian approach. *J. Water Resour. Plan. Manag.* 141 (9), 04015010.
- McKenzie, R., Wegelin, W., 2009. Implementation of pressure management in municipal water supply systems. In: *EYDAP Conference ?Water: The Day after?, Greece*.
- Meseguer, J., Mirats-Tur, J.M., Cembrano, G., Puig, V., Quevedo, J., Pérez, R., Sanz, G., Ibarra, D., 2014. A decision support system for on-line leakage localization. *Environ. Model. Softw.* 60, 331–345.
- Misiunas, D., Vitkovský, J., Olsson, G., Simpson, A., Lambert, M., 2005. Pipeline break detection using pressure transient monitoring. *J. Water Resour. Plan. Manag.* 131 (4), 316–325.
- Page, E.S., 1954. Continuous inspection schemes. *Biometrika* 41 (1/2), 100–115.
- Puust, R., Kapelan, Z., Savic, D.A., Koppel, T., 2010. A review of methods for leakage management in pipe networks. *Urban Water J.* 7 (1), 25–45.
- Rajani, B., Kleiner, Y., 2010. Fatigue failure of large-diameter cast iron mains. In: *12th Annual Conference on Water Distribution Systems Analysis*. WDSA ASCE, pp. 191–200.
- Ratanamahatana, C.A., Keogh, E., 2005. Multimedia retrieval using time series representation and relevance feedback. In: *International Conference on Asian Digital Libraries*. Springer, pp. 400–405.
- Rathnayaka, S., Keller, R., Kodikara, J., Chik, L., 2016. Numerical simulation of pressure transients in water supply networks as applicable to critical water pipe asset management. *J. Water Resour. Plan. Manag.* 142 (6), 04016006.
- Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65.
- Schijve, J., 2009. *Fatigue as a Phenomenon in the Material*. Springer Netherlands, Dordrecht, pp. 13–58.
- Srirangarajan, S., Allen, M., Preis, A., Iqbal, M., Lim, H.B., Whittle, A.J., 2013. Wavelet-based burst event detection and localization in water distribution systems. *Journal of Signal Processing Systems* 72 (1), 1–16.
- Starzewski, D., Collins, R., Boxall, J., 2015. Occurrence of transients in water distribution networks. *Procedia Engineering* vol. 119, 1473–1482. *Computing and Control for the Water Industry (CCWI2015) Sharing the Best Practice in Water Management*.
- Tran, D., Wagner, M., 2002. Fuzzy c-means clustering-based speaker verification. *AFSS International Conference on Fuzzy Systems*. Springer, pp. 318–324.
- Trimble Water Website, 2018. <https://www.trimblewater.com>.
- USEPA, 2016. State of research on high-priority distribution systems issues - EPA. Technical report. U.S. Environmental Protection Agency.
- Visenti, 2018. *Keeping an Intelligent Eye on Your Assets*. <https://www.visenti.com/>.
- Wood, D.J., Lingireddy, S., Boulos, P.F., Karney, B.W., McPherson, D.L., 2005. Numerical methods for modeling transient flow in distribution systems. *J. Am. Water Work. Assoc.* 97 (7), 104–115.
- Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.H., Steinbach, M., Hand, D.J., Steinberg, D., 2008. Top 10 algorithms in data mining. *Knowl. Inf. Syst.* 14 (1), 1–37.
- Wu, Y., Liu, S., Wu, X., Liu, Y., Guan, Y., 2016. Burst detection in district metering areas using a data driven clustering algorithm. *Water Res.* 100, 28–37.
- Wylie, E.B., Streeter, V.L., Suo, L., 1993. *Fluid Transients in Systems*, first ed. Prentice Hall, Englewood Cliffs, NJ.

- Xiong, Y., Yeung, D.Y., 2002. Mixtures of arma models for model-based time series clustering. In: *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on.* IEEE, pp. 717–720.
- Xu, X., Karney, B., 2017. An overview of transient fault detection techniques. In: *Modeling and Monitoring of Pipelines and Networks*. Springer, pp. 13–37.
- Yu, M., Chen, W., Kania, R., Boven, G.V., Been, J., 2016. Crack propagation of pipeline steel exposed to a near-neutral pH environment under variable pressure fluctuations. *Int. J. Fatigue* 82 (Part 3), 658–666.
- Zhao, J., Chen, W., Chevli, K., Been, J., Boven, G.V., Keane, S., Kania, R., 2017. Effect of pressure sampling methods on pipeline integrity analysis. *J. Pipeline Syst. Eng. Pract.* 8 (4), 04017016.
- Zhao, J., Chevli, K., Yu, M., Been, J., Keane, S., Boven, G.V., Kania, R., Chen, W., 2016. Statistical analysis on underload-type pipeline spectra. *J. Pipeline Syst. Eng. Pract.* 7 (4), 04016007.